A general formulation of the cross-nested logit model

Michel Bierlaire, EPFL

Conference paper STRC 2001
Session: Choices

STRC | 1st Swiss Transport Research Conference
Monte Verità / Ascona, March 1.-3. 2001

A general formulation of the cross-nested logit model

Michel Bierlaire, Dpt of Mathematics, EPFL, Lausanne

Phone:	+41.21.693.25.37
Fax:	+41.21.693.55.70
Email:	michel.bierlaire@epfl.ch

Abstract

The cross-nested logit (CNL) model has been mentioned for the first time by Vovsha (1997) in the context of a mode choice survey in Israel. Actually, it is almost identical to the Ordered GEV model proposed by Small (1987). This model, member of the GEV family (McFadden, 1978), is appealing for its ability to capture a wide variety of correlation structures. Papola (2000) has shown that a specific CNL model can be derived for any given homoschedastic variance-covariance matrix. Therefore, the CNL model, with a closed form formulation derived from the GEV model, becomes a serious competitor for the probit model.

In this paper, we develop on the general formulation of the Cross Nested Logit model proposed by Ben-Akiva and Bierlaire (1999). We show that the formulations by Small (1987), Vovsha (1997) and Papola (2000) are particular cases of our general formulation. We also provide some new insights in the CNL model based on theoretical analysis and empirical tests.

We show that one of the conditions imposed by Small (1987), Vovsha (1997) and Papola (2000) is not necessary for the CNL model to be consistent with random utility theory. This condition imposes that the sum of the parameters capturing the degree of membership of an alternative to a nest is equal to one. We prove that the CNL model is a GEV model without using that condition. It is actually a normalization condition, important for parameters identification, but not for model formulation. Finally, we use a simplistic model to illustrate the role of the normalization condition. We propose a variant of the condition proposed in the literature which is slightly more general.

Keywords: logit model, cross-nested model, GEV model, random utility, transportation demand

Swiss Transport Research Conference – STRC 2001 – Monte Verita

1 Introduction

Discrete choice models play a major role in many fields involving a human dimension, including transportation demand analysis (the recent Nobel prize to D. McFadden is a perfect illustration of that statement). Their nice and strong theoretical properties, and their flexibility to capture various situations, provide a vast topic of interest for both researchers and practitioners, that has (by far) not been totally exploited yet. The particular structure of transportation related choice situations are not always fully consistent with the underlying modeling theory (Ben-Akiva and Bierlaire, 1999), requiring to enhance and adapt existing models. The theory on GEV models have been introduced by McFadden (1978). It provides a tremendous potential, as it defines a whole family of models, consistent with random utility theory. It appears that only a few members of this family have been exploited so far. Among them, the cross-nested logit (CNL) model has been used by Vovsha (1997) in the context of a mode choice survey in Israel. Actually, Vovsha's model is almost identical to the Ordered GEV model proposed by Small (1987). This model is appealing for its ability to capture a wide variety of correlation structures. Papola (2000) has shown that a specific CNL model can be obtained for any given homoschedastic variance-covariance matrix. Therefore, the CNL model, with a closed form formulation derived from the GEV model, becomes a serious competitor for the probit model. It has been shown to be specifically appropriate for route choice applications (Vovsha and Bekhor, 1998), where topological correlations cannot be captured correctly by the multinomial and the nested logit models. Unfortunately, no satisfactory procedure has been proposed for the estimation of the model parameters.

The remaining of this section introduces the GEV model and presents various formulations of the Cross-Nested Logit model. In Section 2, we analyze the most general formulation. We prove that it is consistent with the GEV model family.

1.1 The GEV model

The Generalized Extreme Value (GEV) model has been derived from the random utility model by McFadden (1978). This general model consists of a large family of models that include the Multinomial Logit and the Nested Logit models. The probability of choosing alternative i within the choice set C of a given choice maker is

$$P(i|C) = \frac{y_i \frac{\partial G}{\partial y_i}(y_1, \ldots, y_J)}{\mu G(y_1, \ldots, y_J)} \tag{1}$$

where J is the number of available alternatives, $y_i = e^{V_i}$, V_i is the deterministic part of the utility function associated to alternative i, and G is a non-negative differentiable function defined on $\mathbb{R}^J_+$ with the following properties:

1. G is homogeneous of degree $\mu > 0$, that is $G(\alpha y) = \alpha^\mu G(y)$,
2. $\lim_{y_i \to +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,
3. the kth partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even that is, for any distinct indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \ldots \partial x_{i_k}}(x) \leq 0, \ \forall x \in \mathbb{R}^J_+. \tag{2}$$

Note that the homogeneity of G and Euler's theorem give

$$P(i|C) = \frac{e^{V_i + \ln G_i(\ldots)}}{\sum_{j=1}^J e^{V_j + \ln G_j(\ldots)}}, \tag{3}$$

where $G_i = \frac{\partial G}{\partial y_i}$.

It is well known that the Multinomial Logit and the Nested Logit models are instances of this model family. We present now several formulations of the Cross-Nested Logit model, derived from the GEV model.

1.2 Formulations of the Cross-Nested Logit model

The limitations of the Nested Logit model has been observed by several authors (Williams, 1977, Forinash and Koppelman, 1993). The requirement of unambiguous assignment of alternatives to nests does not allow to capture mixed interactions across alternatives.

It seems that the first Cross-Nested Logit model has been proposed by Small (1987) in the context of departure time choice. Small's model, called the *Ordered GEV model*, is based on the following function:

$$G(y_1, \ldots, y_j) = \sum_{r=1}^{J+M} \left(\sum_{B_r} w_{r-j} y_j^{1/\rho_r} \right)^{\rho_r}, \tag{4}$$

where M is a positive integer, ρ_r and w_m are constants satisfying $0 < \rho_r \leq 1$, $w_m \geq 0$ and

$$\sum_{m=0}^{M} w_m = 1. \tag{5}$$

The B_r are overlapping subsets of alternatives:

$$B_r = \{j \in \{1, \ldots, J\} | r - M \leq j \leq r\}. \tag{6}$$

Vovsha (1997) applies the Cross-Nested Logit model to a mode choice application, where the "park&ride" alternative is allowed to belong to the "composite auto" and the "composite transit" nests. Vovsha derives the Cross-Nested Logit from the GEV model with the generating function:

$$G(y_1, \ldots, y_J) = \sum_m \left(\sum_{j \in C} \alpha_{jm} y_j \right)^{\mu} \tag{7}$$

where m is the nest index, and α_{jm} are model parameters such that

$$0 \leq \alpha_{jm} \leq 1 \quad \forall j, m, \tag{8}$$

and

$$\sum_m \alpha_{jm} > 0 \quad \forall i. \tag{9}$$

Vovsha (1997) imposes also that

$$\sum_m \alpha_{im}^{\mu} = 1 \quad \forall i. \tag{10}$$

Ben-Akiva and Bierlaire (1999) mention the CNL as an instance of a GEV model, and based on the following generating function:

$$G(y_1,\ldots,y_J)=\sum_m\left(\sum_{j\in C}\alpha_{jm}y_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}}. \tag{11}$$

A similar formulation is used by Papola (2000), based on the following generating function:

$$G(y_1,\ldots,y_J)=\sum_k\left(\sum_{j\in C_k}\alpha_{ik}^{\theta_0/\theta_k}e^{V_i/\theta_k}\right)^{\frac{\theta_k}{\theta_0}}, \tag{12}$$

with $0\leq\theta_k\leq\theta_0$. Papola imposes also that

$$\sum_k\alpha_{ik}=1\quad\forall i. \tag{13}$$

2 Theoretical analysis

Among these formulations, (11) is the most general. Indeed, Vovsha's and Small's formulations are specific cases of (11). We obtain Small's formulation (4) with $\mu=1$ and $\mu_m=1/\rho_m$. Vovsha's formulation (7) is obtained from (11) with $\mu_m=1$ for all m.

Papola's model (12) is equivalent to (11), with $\mu=1/\theta_0$, $\mu_m=1/\theta_m$ and $\alpha_{jm}=\alpha_{jm}^{\theta_0/\theta_m}$. However, Papola's constraint (13) is not required by our formulation. Note that Small (5) and Vovsha (10) impose the same constraint.

The following theorem shows that (11) is indeed a GEV generating function.

Theorem 1 *The following conditions are sufficient for (11) to define a GEV generating function:*

1. $\alpha_{jm}\geq 0$, $\forall j,m$,
2. $\sum_m\alpha_{jm}>0$, $\forall j$,
3. $\mu>0$,
4. $\mu_m>0$, $\forall m$,
5. $\mu\leq\mu_m$, $\forall m$.

Proof. We show that, under these assumptions, (11) verifies the four properties of GEV generating functions.

1. G is obviously non negative, if $y \in \mathbb{R}^n_+$.
2. G is homogeneous of degree μ. Indeed,

$$\begin{aligned} G(\beta y) &= \sum_m \left(\sum_{j \in C} \alpha_{jm} \beta^{\mu_m} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\ &= \sum_m \left(\beta^{\mu_m} \sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\ &= \sum_m \beta^{\mu} \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\ &= \beta^{\mu} \sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\ &= \beta^{\mu} G(y). \end{aligned}$$

3. The limit properties hold from assumption 2, that guarantees that there is at least one non zero coefficient α_{jm} for each alternative j.

$$\begin{aligned} \lim_{y_i \to \infty} G(y_1, \ldots, y_J) &= \lim_{y_i \to \infty} \textstyle\sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\ &= \textstyle\sum_m \left(\lim_{y_i \to \infty} \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \right) \\ &= \infty \end{aligned} \tag{14}$$

4. The condition for the sign of the derivatives is obtained from (21) in Lemma 2 (see Section 6). We distinguish three cases, considering only $y \geq 0$.

   (a) If $k = 1$, we have

$$\partial G(y) / \partial y_j = \sum_m \left(\mu \alpha_{jm} y_j^{\mu_m - 1} A_m^{\frac{\mu - \mu_m}{\mu_m}} \right) \geq 0. \tag{15}$$

   (b) If $k > 1$ and $\mu = \mu_m$, we have

$$\partial^k G(y) / \partial y_{i_1} \ldots \partial y_{i_k} = 0. \tag{16}$$

Indeed,

$$\prod_{n=0}^{k-1}\left(\frac{\mu}{\mu_m}-n\right) \tag{17}$$

contains a zero factor when $n=1$.

(c) If $k>1$ and $\mu<\mu_m$, the sign of (21) is entirely determined by the sign of (17). For $n>0$, we have $\frac{\mu}{\mu_m}-n<0$ (assumption 5). Therefore, there are $k-1$ negative and one positive factors in the product. We obtain that

$$\prod_{n=0}^{k-1}\left(\frac{\mu}{\mu_m}-n\right)\begin{cases}\geq 0 & \text{if } k \text{ is odd}\\ \leq 0 & \text{if } k \text{ is even}\end{cases} \tag{18}$$

Therefore, in any case, we have

$$\partial^k G(y)/\partial y_{i_1}\ldots\partial y_{i_k}\begin{cases}\geq 0 & \text{if } k \text{ is odd}\\ \leq 0 & \text{if } k \text{ is even}\end{cases} \tag{19}$$

□

This theorem motivates the absence in our formulation of a constraint similar to (5), (10) and (13). Indeed, such a condition is not required to obtain a valid GEV model and, consequently, to be consistent with discrete choice theory. Instead, they are used to enable parameter estimability. Indeed, it is impossible to estimate all parameters of the GEV model, exactly as it is impossible to estimate all Alternative Specific constants in a MNL or NL model (see Bierlaire, Lotan and Toint, 1995).

3 Estimation procedure

The estimation procedures proposed by Small (1987) and Vovsha (1997) are based on heuristics. Small reduces the number of free parameters by imposing arbitrary restrictions on the parameters: $w_m=\frac{1}{M+1}$, $\forall m$, and $\rho_r=\rho$, $\forall r$. Vovsha proposes a complicated heuristic, where each observation is artificially substituted with n observations (Vovsha proposes n =100).

We prefer to use a pure maximum likelihood procedure. However, the problem of estimability remains open. A detailed theoretical analysis of model overspecification, similar to what Bierlaire et al. (1995) has done for the Alternative Specific

Constants in the Nested Logit model, is currently ongoing. Meanwhile, we have developed an optimization process which is able to handle overspecified models.

A new model estimation package called Biogeme (BIerlaire's Optimization routines for GEv Model Estimation) has been developed. It is designed to estimate a wide variety of discrete choice models. Actually, any model out of the GEV model family can be estimated. Moreover, non linear utility functions can be handled. In particular, a specific scale parameter can be associated with different groups in the sample.

The optimization algorithm is based on a quasi-Newton BFGS method in a trust-region framework (Conn, Gould and Toint, 2000). The trust-region subproblem is solved with a conjugate gradient method, which does not require the solution of the Newton equations. These equations do not have a solution when the approximation of the hessian matrix is singular (as it is the case with overspecified models). Instead, the conjugate gradient involves only matrix-vector computations. The Newton method is known to be very slow when the objective is not strictly convex at the solution. Also, it stops when a local maximum of the log-likelihood function has been reached. The search for a global maximum is out of the scope of the method. Despite these shortcomings, this optimization algorithm is sufficiently robust to empirically analyze the structure of the CNL model. An efficient estimation procedure will be derived later on, when the structural analysis will be complete.

4 Preliminary empirical analysis

We have performed a preliminary empirical analysis of the model estimability. Our main objective is to assess the relevance of conditions (5), (10) and (13) mentioned in the literature.

We analyze a trivial model, with 3 alternatives. The utility functions are just composed of the ASCs. The sample contains 3 observations, each one corresponding to a different alternative. It is well know that the maximum likelihood estimator for a MNL model is obtained with all ASC being equal (say to zero). The associated log-likelihood is $-3\ln 3 \approx -3.29584$. Clearly, using a CNL model for such a trivial model cannot improve the log-likelihood. But we see that the same log-likelihood can be obtained with various combinations of parameters values.

	1	2	3	4
asc1	0.0	0.0	0.0	0.0
asc2	0.0	0.0	0.0	-0.45
asc3	0.0	0.0	0.0	-0.69
μ	1.0	1.0	1.0	1.0
μ_1	1.0	1.0	1.0	1.0
μ_2	1.0	1.0	1.0	1.0
α_{11}	0.5	0.0	1.0	1.0
α_{21}	0.5	1.0	1.0	1.0
α_{31}	0.5	1.0	1.0	1.0
α_{12}	1.0	2.0	1.0	1.0
α_{22}	1.0	1.0	1.0	2.0
α_{23}	1.0	1.0	1.0	3.0

Table 1: Optimal parameters for the trivial model

We conjecture that if

$$\sum_m \alpha_{im}^{\mu/\mu_m} = K, \quad \forall i, K > 0, \tag{20}$$

then the interpretation of the ASC is consistent with random utility theory. Note that condition (5) by Small, condition (10) by Vovsha and condition (13) by Papola are equivalent to (20), with $K = 1$. If (20) is not verified, we still have a valid model with a perfect predictive capability. However, the usual interpretation of the ASC parameters is not relevant anymore.

We illustrate that conjecture in Table 1, which contains sets of maximum likelihood estimators of the parameters. The parameters contained in each column provide the exact same log-likelihood ($-3\ln 3$). Condition (20) is verified for columns 1 to 3. We observe that the optimal value of the ASC are the same as the MNL. In column 4, where the condition is not verified, the values of the ASC are completely different. We note also that the α parameter may take values larger than 1, and that the value of K in(20) is not necessarily 1 (K=1 in column 1, and $K = 2$ in columns 2 and 3). We have observed similar behavior with a non trivial model on a combined RP/SP data set for mode choice in Switzerland (Bierlaire, Axhausen and Abbay, 2001). In that case, it seems that the "regular" parameters (for cost or time) are independent from the validity of 20, while the

ASCs are clearly affected.

We have shown that (20) is not necessary for the CNL to be part of the GEV model family. Therefore, it may be adequate to ignore it. We suggest that condition (20) is violated for estimation purposes, as it provides more degrees of freedom to the algorithms and avoid to handle linear constraints. But once an estimated model is obtained, an equivalent model verifying (20) must be derived, in order to provide values of the parameters that can be analyzed in a usual way. We also note that the constant K in (20) does not necessarily need to be one. Formal proofs of these statements must of course be provided. Such proofs are straightforward for the trivial model used for empirical analysis, but not interesting. The proof for the general case is much more difficult. This research is currently ongoing.

5 Conclusion and perspectives

In this paper, we have provided and analyzed a general formulation of the cross-nested logit model. It generalizes existing formulations in the literature. We have proved that, under mild conditions, the formulation is consistent with GEV models.

Also, we have performed preliminary analysis of a condition on the parameters imposed on formulations published in the literature. We have proved that this condition is not necessary for the theoretical properties of the model. But it is important for practical properties, related to parameter estimability, and model comparison. We have identified an extension of this condition that is important in order to correctly interpret the estimated parameters. The condition is therefore desirable for practical purposes, but not necessary to have a model with prediction capabilities.

The CNL model is appealing to capture complex situations where correlations cannot be handled by the Nested Logit model. Even with few alternative and nests, the use of a CNL instead of a NL model may significantly improve the estimated model (Bierlaire et al., 2001). The price to pay is the difficulty to estimate the model parameters. We have adopted a robust method with does not require the model to be fully estimable in order to find an optimal value of the parameters. However, because the second derivative matrix of the log-likelihood function is singular at the solution, we cannot compute a reliable estimation of the variance covariance matrix associated with the estimates. Also, the estimation algorithm, based on a quasi-Newton approach, can be very slow.

An formal analysis of the sources of overspecification, similar to what Bierlaire et al. (1995) have made for the Nested Logit Model, is required in order to obtain an efficient estimation process, and a good estimates of the variance-covariance matrix of the estimated parameters.

6 Annex: Lemma

The following Lemma has been proven by Nicolas Antille, who is gratefully acknowledged.

Lemma 2 *Let $i_1,\ldots,i_k$ be k different indices ($k > 0$) arbitrarily chosen within $\{1,\ldots,J\}$. If G is defined by (11), then $\partial^k G(y)/\partial y_{i_1}\ldots\partial y_{i_k} =$*

$$\sum_m \left(\mu_m^k \prod_{n=i_1}^{i_k} \left(\alpha_{nm} y_n^{\mu_m - 1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) A_m^{\frac{\mu - k\mu_m}{\mu_m}} \right) \tag{21}$$

where

$$A_m = \sum_{j\in C} \alpha_{jm} y_j^{\mu_m}. \tag{22}$$

Proof. The proof is by induction. We have

$$\begin{aligned} \frac{\partial^k G(y)}{\partial y_{i_1}} &= \sum_m \left(\frac{\mu}{\mu_m} A_m^{\frac{\mu-\mu_m}{\mu_m}} \mu_m \alpha_{i_1 m} y_{i_1}^{\mu_m - 1} \right) \\ &= \sum_m \left(\mu_m \alpha_{i_1 m} y_{i_1}^{\mu_m - 1} \frac{\mu}{\mu_m} A_m^{\frac{\mu-\mu_m}{\mu_m}} \right) \end{aligned}$$

proving the result for $k = 1$.

Assuming now that the result is verified for k, we have $\partial^{k+1} G(y)/\partial y_{i_1}\ldots\partial y_{i_{k+1}} =$

$$\begin{aligned} & \frac{\partial}{\partial y_{i_{k+1}}} \frac{\partial^k G(y)}{\partial y_{i_1}\ldots\partial y_{i_k}} \\ = \; & \frac{\partial}{\partial y_{i_{k+1}}} \sum_m \left(\mu_m^k \prod_{n=i_1}^{i_k} \left(\alpha_{nm} y_n^{\mu_m-1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n\right) A_m^{\frac{\mu-k\mu_m}{\mu_m}} \right) \\ = \; & \sum_m \mu_m^k \prod_{n=i_1}^{i_k} \left(\alpha_{nm} y_n^{\mu_m-1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n\right) \left(\frac{\mu}{\mu_m} - k\right) A_m^{\frac{\mu-k\mu_m-\mu_m}{\mu_m}} \alpha_{i_{k+1}} \mu_m y_{i_{k+1}}^{\mu_m-1} \\ = \; & \sum_m \left(\mu_m^{k+1} \prod_{n=i_1}^{i_{k+1}} \left(\alpha_{nm} y_n^{\mu_m-1}\right) \prod_{n=0}^{k} \left(\frac{\mu}{\mu_m} - n\right) A_m^{\frac{\mu-(k+1)\mu_m}{\mu_m}} \right). \end{aligned}$$

That concludes the proof. $\square$

7 Annex: Derivatives

We provide here the derivatives of the log-likelihood function for GEV models in general, and for the Cross-Nested Logit model in particular. Given a sample of observations, the log-likelihood of the sample is

$$\mathcal{L} = \sum_{n \in \text{sample}} \ln P(i_n | C_n), \tag{23}$$

where i_n is the alternative actually chosen by individual n, C_n is the choice set, and

$$\begin{array}{rcl} \ln P(i_n|C) & = & \lambda_n V_{i_n} \\ & + & \ln G_{i_n}(e^{\lambda_n V_1}, \ldots, e^{\lambda_n V_J}) \\ & - & \ln \left(\sum_j e^{\lambda_n V_j} G_j(e^{\lambda_n V_1}, \ldots, e^{\lambda_n V_J}) \right) \end{array} \tag{24}$$

where $G_i = \partial G / \partial y_i$ and λ_n is a scale parameter associated to individual n. This parameter allows to estimate models with heterogeneous samples, without using complicated nested structures.

If β_k is a parameter appearing in the utility functions $V_1, \ldots, V_J$, we have

$$\begin{array}{rcl} \frac{\partial}{\partial \beta_k} \ln P & = & \lambda \frac{\partial V_i}{\partial \beta_k} \\ & + & \frac{1}{G_i} \sum_{j=1}^{J} \frac{\partial G_i}{\partial x_j} e^{\lambda V_j} \lambda \frac{\partial V_j}{\partial \beta_k} \\ & - & \frac{1}{\Delta} \sum_j e^{\lambda V_j} \left(\lambda \frac{\partial V_j}{\partial \beta_k} G_j + \sum_{n=1}^{J} \frac{\partial G_j}{\partial x_n} e^{\lambda V_n} \lambda \frac{\partial V_n}{\partial \beta_k} \right) \end{array} \tag{25}$$

where

$$\Delta = \sum_j e^{\lambda V_j} G_j \tag{26}$$

Note that we do not assume here that the V_j are linear-in-parameters, so that $\partial V_j / \partial \beta_k$ is not necessarily a constant.

The derivatives with respect to model parameters $\alpha_{k\ell}$ are given by

$$\frac{\partial}{\partial \alpha_{k\ell}} \ln P = \frac{1}{G_i} \frac{\partial G_i}{\partial \alpha_{k\ell}} - \frac{1}{\Delta} \sum_j e^{\lambda V_j} \frac{\partial G_j}{\partial \alpha_{k\ell}} \tag{27}$$

The derivatives with respect to model parameters μ_k are given by

$$\frac{\partial}{\partial \mu_k} \ln P = \frac{1}{G_i} \frac{\partial G_i}{\partial \mu_k} - \frac{1}{\Delta} \sum_j e^{\lambda V_j} \frac{\partial G_j}{\partial \mu_k} \tag{28}$$

The derivative with respect to model parameter μ is given by

$$\frac{\partial}{\partial \mu} \ln P = \frac{1}{G_i} \frac{\partial G_i}{\partial \mu} - \frac{1}{\Delta} \sum_j e^{\lambda V_j} \frac{\partial G_j}{\partial \mu} \tag{29}$$

The derivative with respect to the scale parameter λ is given by

$$\frac{\partial}{\partial \lambda} \ln P = V_i + \frac{1}{G_i}\frac{\partial G_i}{\partial \lambda} - \frac{1}{\Delta}\frac{\partial \Delta}{\partial \lambda} \tag{30}$$

where

$$\frac{\partial G_i}{\partial \lambda} = \sum_j V_j e^{\lambda V_j} \frac{\partial G_i}{\partial x_j} \tag{31}$$

and

$$\frac{\partial \Delta}{\partial \lambda} = \sum_j \left(V_j e^{\lambda V_j} G_j + e^{\lambda V_j} \frac{\partial G_j}{\partial \lambda} \right) \tag{32}$$

Finally, we provide the first and second derivatives of (11) with respect to every parameter. The first derivative with respect to a variable x_i is given by

$$G_i = \frac{\partial G}{\partial x_i} = \mu \sum_m \alpha_{im} x_i^{\mu_m - 1} \left(\sum_j \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1} \tag{33}$$

The first derivative with respect to the μ parameter is

$$\frac{\partial G}{\partial \mu} = \sum_m \frac{1}{\mu_m} y_m^{\frac{\mu}{\mu_m}} \ln(y_m) \tag{34}$$

where

$$y_m = \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \tag{35}$$

The first derivative with respect to the nest parameter μ_m is

$$\frac{\partial G}{\partial \mu_m} = \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \ln(x_j) \right) - \frac{\mu}{\mu_m^2} y_m^{\frac{\mu}{\mu_m}} \ln(y_m) \tag{36}$$

and with respect to the α parameter is

$$\frac{\partial G}{\partial \alpha_{im}} = \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} x_i^{\mu_k} \tag{37}$$

where

$$y_m = \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \tag{38}$$

We now provide the second derivative with respect to x_i and x_j. If $i = j$, we have

$$\frac{\partial^2 G}{\partial x_i^2} = \frac{\partial G_i}{\partial x_i} = \sum_m \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 2} \alpha_{im} \mu_m x_i^{\mu_m - 2} \left(\left(\frac{\mu}{\mu_m} - 1 \right) \alpha_{im} \mu_m x_i^{\mu_m} + y_m(\mu_m - 1) \right) \tag{39}$$

and if $i \neq j$, we have

$$\frac{\partial^2 G}{\partial x_i \partial x_j} = \frac{\partial G_i}{\partial x_j} = \sum_m \mu_m \mu (\frac{\mu}{\mu_m} - 1) \alpha_{im} \alpha_{jm} y_m^{\frac{\mu}{\mu_m}-2} x_i^{\mu_m - 1} x_j^{\mu_m - 1} \tag{40}$$

where

$$y_m = \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \tag{41}$$

The second derivative with respect to x_i and μ is

$$\frac{\partial^2 G}{\partial x_i \partial \mu} = \frac{\partial G_i}{\partial \mu} = \sum_m y_m^{\frac{\mu}{\mu_m}-1} \alpha_{im} x_i^{\mu_m - 1} (1 + \frac{\mu}{\mu_m} \ln(y_m)) \tag{42}$$

where

$$y_m = \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \tag{43}$$

The second derivative with respect to x_i and μ_m is

$$\begin{aligned} \tfrac{\partial^2 G}{\partial x_i \partial \mu_m} &= \tfrac{\partial G_i}{\partial \mu_m} \\ &= -\tfrac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m}-1} \alpha_{im} x_i^{\mu_m - 1} - \tfrac{\mu^2}{\mu_m^2} y_m^{\frac{\mu}{\mu_m}-1} \ln(y_m) \alpha_{im} x_i^{\mu_m - 1} \\ &+ \tfrac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m}-1} \alpha_{im} x_i^{\mu_m - 1} + \mu y_m^{\frac{\mu}{\mu_m}-1} \alpha_{im} x_i^{\mu_m - 1} \ln(x_i). \end{aligned} \tag{44}$$

The second derivative with respect to x_i and α_{ik} is

$$\frac{\partial^2 G}{\partial x_i \partial \alpha_{ik}} = \mu x_i^{\mu_k - 1} y_k^{\frac{\mu}{\mu_k}-1} \left(1 + \alpha_{ik} (\frac{\mu}{\mu_k} - 1) y_k^{-1} x_i^{\mu_k}\right) \tag{45}$$

and with respect to x_i and α_{jk} $(i \neq j)$ is

$$\frac{\partial^2 G}{\partial x_i \partial \alpha_{jk}} = \mu \alpha_{ik} x_i^{\mu_k - 1} (\frac{\mu}{\mu_k} - 1) y_k^{\frac{\mu}{\mu_k}-2} x_j^{\mu_k} \tag{46}$$

where

$$y_m = \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \tag{47}$$

References

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions, *in* R. Hall (ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.

Bierlaire, M., Axhausen, K. and Abbay, G. (2001). Acceptance of model innovation: the case of the swissmetro, *Proceedings of the 1st Swiss Transport Research Conference*.

Bierlaire, M., Lotan, T. and Toint, P. L. (1995). On the overspecification of multinomial and nested logit models due to alternative specific constants, *Technical Report 95/12*, Department of Mathematics, FUNDP, Namur, Belgium.

Conn, A., Gould, N. and Toint, P. (2000). *Trust region methods*, MPS–SIAM Series on Optimization, SIAM.

Forinash, C. V. and Koppelman, F. S. (1993). Application and interpretation of nested logit models of intercity mode choice, *Transportation Research Record Issue* **1413**: 98–106.

McFadden, D. (1978). Modelling the choice of residential location, *in* A. K. et al. (ed.), *Spatial interaction theory and residential location*, North-Holland, Amsterdam, pp. 75–96.

Papola, A. (2000). Some development of the cross-nested logit model, *Proceedings of the 9th IATBR Conference*.

Small, K. (1987). A discrete choice model for ordered alternatives, *Econometrica* **55**(2): 409–424.

Vovsha, P. (1997). Cross-nested logit model: an application to mode choice in the Tel-Aviv metropolitan area, Transportation Research Board, 76th Annual Meeting, Washington DC. Paper #970387.

Vovsha, P. and Bekhor, S. (1998). The link-nested logit model of route choice: overcoming the route overlapping problem, *Transportation Research Record* **1645**: 133–142.

Williams, H. (1977). On the formation of travel demand models and economic measures of user benefit, *Environment and Planning* **9A**: 285–344.